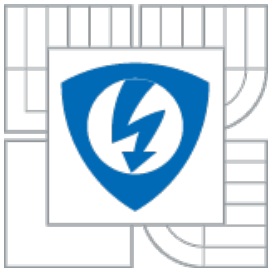
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV AUTOMATIZACE A MĚŘICÍ TECHNIKY

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF CONTROL AND INSTRUMENTATION

ROZPOZNÁNÍ JEDNOTLIVÝCH PÍSMEN VE ZVUKOVÉM ZÁZNAMU S VYUŽITÍM SOM

CHARACTER RECOGNITION IN THE SOUNDTRACK WITH SOM

BAKALÁŘSKÁ PRÁCE
BACHELOR´S THESIS

AUTOR PRÁCE — JAN MALÁSEK
AUTHOR

VEDOUCÍ PRÁCE — ING. JAN POHL
SUPERVISOR

BRNO 2010

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky a komunikačních technologií

Ústav automatizace a měřicí techniky

Bakalářská práce

bakalářský studijní obor
Automatizační a měřicí technika

Student: Jan Malásek ***ID:*** 74870
Ročník: 3 ***Akademický rok:*** 2009/2010

NÁZEV TÉMATU:

Rozpoznání jednotlivých písmen ve zvukovém záznamu s využitím SOM

POKYNY PRO VYPRACOVÁNÍ:

1. Seznamte se s problematikou zpracování řeči pomocí umělých neuronových sítí.
2. Pro popis řečového signálu využijte FFT.
3. Vstupem navrženého algoritmu bude záznam jednotlivých vybraných písmen.
4. Výstupem algoritmu bude textový zápis rozpoznaného písmene.
5. Algoritmus bude pro klasifikaci využívat SOM.
6. Práce bude obsahovat zhodnocení dosažených výsledků.

DOPORUČENÁ LITERATURA:

Katagiri, S: Handbook of Neural Networks for Speech Processing, Artech House Inc, 200, ISBN 0890069549
Šíma, J., Neurda, R., Teoretické otázky neuronových sítí. Praha: MATFYZPRESS, 1996. ISBN 80-85863-18-9
Psutka, J., Komunikace s počítačem mluvenou řečí, ACADEMIA Praha 1995, ISBN 80-200-0203-0

Termín zadání: 8.2.2010 ***Termín odevzdání:*** 31.5.2010

Vedoucí práce: Ing. Jan Pohl

prof. Ing. Pavel Jura, CSc.
Předseda oborové rady

UPOZORNĚNÍ:

ANOTACE

Bakalářská práce popisuje historické pozadí vývoje neuronových sítí a jejich použití při procesu rozpoznání řeči a uvádí do problematiky práce a učení neuronových sítí. Představuje tři vybrané systémy pro rozpoznání řečového signálu včetně vyhodnocení jejich úspěšnosti v experimentech, výhod a nevýhod. Zabývá se charakteristikou lidské řeči a systémy na její rozpoznávání. Nabízí pohled na spektra signálů různých typů hlásek a dává návod k programování neuronových sítí v prostředí MATLAB.

KLÍČOVÁ SLOVA:

Rozpoznávání řeči, neuronové sítě, forward-backward algoritmus, skryté Markovovy modely, foném, učení, SOM, frekvenční spektrum.

ABSTRACT

This bachelor´s thesis describes a history of neural networks evolution and their using in speech recognition systems and shows problems with working and learning neural networks. It presents three chosen systems for speech recognition including their evaluation in experiments, their advantages and disadvantages. It is also about human speech characteristics and systems of its recognition. The last part is focused on frequency spectrums of different types of vowels and gives instructions for programming neural networks using MATLAB.

KEYWORDS:

Speech recognition, neural networks, forward-backward, Hidden Markov Models, phone, learning, SOM, frequency spectrum.

Bibliografická citace

MALÁSEK, Jan. *Rozpoznání jednotlivých písmen ve zvukovém záznamu s využitím SOM*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2010. 53 s., Vedoucí práce Ing. Jan Pohl.

P r o h l á š e n í

„Prohlašuji, že jsem svoji bakalářskou práci na téma „Rozpoznání jednotlivých písmen ve zvukovém záznamu s využitím SOM“ vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s jejím vytvořením jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.“

V Brně dne: Podpis:

P o d ě k o v á n í

Děkuji tímto Ing. Janu Pohlovi za odborné vedení mé bakalářské práce.

V Brně dne: Podpis:

OBSAH

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

1. ÚVOD

Řešení problematiky komunikace člověka s počítačem má dlouhou historii. V současné době je jednou z nejvíce se vyvíjejících vědních disciplín zasahující do mnoha oborů lidské činnosti. V oblasti výzkumu je pozornost soustředěna na odstranění problémů s rozpoznáváním slov a slovních spojení, na co největší zvýšení spolehlivosti a možnost širšího uplatnění.

Jedná se o rozsáhlou problematiku zasahující široce do mnoha vědních disciplín. Tato práce se zabývá systémy pro rozpoznání řeči člověka založenými na bázi umělé neuronové sítě.

První část (kap. 2) se věnuje teoretickému popisu neuronových sítí a jejich vlastností. Dále uvádí několik vybraných modelů, které jsou svoji organizační mechanikou a souhrnem základních vlastností vhodné pro aplikace v problematice automatizovaného rozpoznávání řeči. V tomto smyslu je jejich vhodnost posouzena a vzájemně srovnána.

Druhá část (kap. 3) popisuje již existující systémy pro rozpoznání řečových fragmentů. Uvádí principy jejich funkce, vlastnosti a vzájemné srovnání.

Třetí část (kap. 4) se zabývá charakteristikou lidské řeči a podává základní přehled systémů na její rozpoznávání. Poukazuje na rozdíly mezi jednotlivými systémy a jejich problémy.

Závěr práce je věnován návrhu vlastního systému pro rozpoznání řeči v programovacím prostředí MATLAB.

2. KLASICKÉ MODELY NEURONOVÝCH SÍTÍ

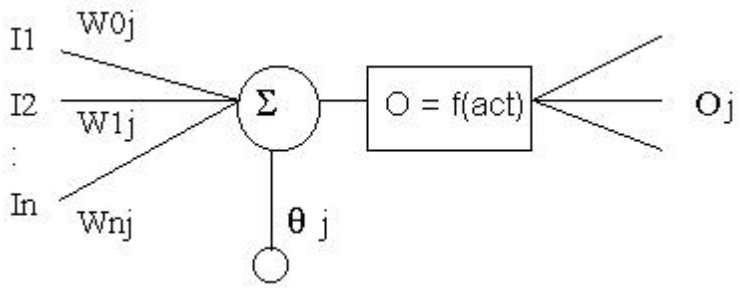
Tato kapitola popisuje stručnou historii vývoje oboru umělých neuronových sítí a jejich inspiraci v biologii. Dále se podrobněji zabývá vybranými modely, které svými schopnostmi umožňují využití v oblasti automatizované-ho rozpoznání lidské řeči.

2.1 HISTORIE NEURONOVÝCH SÍTÍ

Prvotní pokusy

Základy neuronových sítí položili Warren McCulloch a Walter Pitts v roce 1943. Vytvořili matematický popis neuronů a ukázali, že neurony je možno chápat jako logické přepínače fungující dle pravidel Booleovy algebry. Dokázali také, že spojením takovýchto elementárních jednotek do neuronových sítí je možno postavit zařízení schopné provádět libovolné operace výrokového kalkulu.

Abstraktní neuron (Obr. 2.1) je objekt, jenž má jistý počet vstupů $I_1,..,I_n$ a jediný výstup O (to neznamená, že jeho výstup může být vstupem pouze jednoho neuronu, ale pouze to, že pro všechny neurony, jež mají jeho výstup jako vstup, je jeho hodnota stejná) a vnitřní výpočetní funkci f, jejímiž argumenty jsou vstupy a funkční hodnota je výstupem daného neuronu.

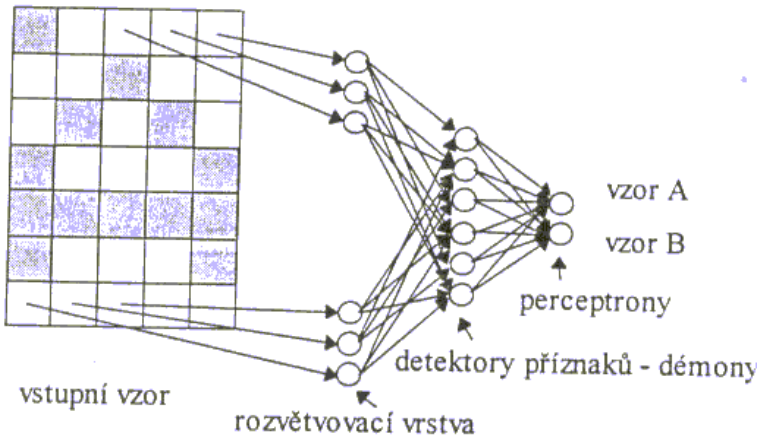


Obr. 2.1 Model neuronu dle návrhu Warrena McCullocha a Waltera Pittse [3]

V roce 1945 Donald Hebb vytvořil návod pro vznik prvního učícího algoritmu. Inspiroval se myšlenkou podmíněných reflexů, které jsou pozorovatelné u živočichů.

Slibně se rozvíjející technologie

K rozvoji neuronových sítí přispěli také psychologové a inženýři. Frank Rosenblatt zobecnil model neuronu v roce 1957 na tzv. perceptron, který počítal s reálnými čísly. Jde o pevnou architekturu jednovrstvé sítě s *n* vstupními a *m* výstupními neurony. Inspiroval se lidským okem.



Obr. 2.2 Architektura perceptronové sítě [3]

Architektura perceptronové sítě (obr 2.2) vycházející z fyziologického modelu je třívrstvá (obecně je perceptronová síť na vrstvy neomezená). Vstupní vrstva funguje jako vyrovnávací. Jejím úkolem je mapování dvourozměrného pole čidel na jednorozměrný vektor procesorových elementů. Tuto druhou vrstvu tvoří detektory příznaků. Každý z nich je náhodně spojen s prvky vstupní vrstvy. Úkolem démonů je detekce specifických příznaků. Poslední, nejdůležitější, vrstva obsahuje rozpoznávače vzorů. Zatímco váhy jsou ve vstupní a druhé vrstvě konstantní, lze váhy na vstupech výstupní vrstvy při trénování nastavovat.

Na základě tohoto výzkumu sestrojil Frank Rosenblatt s Charlesem Wightmanem během let 1957 a 1958 první neuropočítač. Jmenoval se *Mark I Perceptron* a byl navržen pro rozpoznávání znaků. [3]

Dalším typem neuronového prvku, který se velmi podobal perceptronu byl Adaline *(Adaptive Linear Element).* Aktivní dynamika se u tohoto modelu lišila tím, že výstupy sítě byly obecně reálné a jednotlivé Adaline realizovali lineární funkci. Autorem byli Bernard Widrow se svými studenty v roce 1960. V polovině 60. let se zasloužil o vznik první komerční organizace zabývající se stavbou neuropočítačů. [3]

Na přelomu 50. - 60. let zaznamenal úspěch Karl Steinbuch, který vyvinul model binární asociativní sítě. Na rozdíl od paměti klasických počítačů, kdy klíč k vyhledání položky v paměti je adresa, u asociativní paměti dochází k vybavení určité události či informace na základě její částečné znalosti. [3]

Období stagnace

V roce 1969 došlo ke zvratu. Matemetici Minski a Papert vyjádřili meze zobecněných modelů. Dospěli k závěru, že neuronové sítě nemohou nahradit klasické metody, protože s jejich pomocí nelze simulovat všechny logické zákony (pomocí jednovrstvého perceptronu nelze řešit XOR problém). Vysvětlili, že nutná omezení mohou být sice odstraněna konstrukcí vícevrstvého perceptronu, pro který je však nutné nalézt parametry „ručně“, není možné je automaticky předpovědět a nalézt algoritmus učení pro tuto strukturu. Jejich kritika byla pro tuto dobu oprávněná, ale přesto unáhlená.

Velkou chybou bylo, že nepřipustili další vývoj a možný pokrok ve výzkumu. [3]

Období vzestupu

Ačkoliv zájem veřejnosti a financování byly minimální, několik výzkumných pracovníků pokračovalo ve vývoji neuromorfologických výpočetních metod pro řešení problémů s rozeznáváním objektů aj. Během

tohoto období bylo vytvořeno několik paradigmat, které navázaly na předchozí výzkum v oblasti sítí a posunuly jej vpřed.

V roce 1974 vyvinul Paul Verbos učící metodu zpětného šíření *(back-propagation learning method)*. Dnes je tato metoda pravděpodobně jedna z nejznámějších a nejrozšířenějších neuronových sítí dneška. [4]

V roce 1983 došlo v USA k novým investicím do výzkumu neuronových sítí. Vznikla DARPA *(Defence Advance Research Project Association)*. [3]

Vzestup popularity

V roce 1987 se v San Diegu konala první větší konference s výhradním zaměřením na neuronové sítě, jejímž výsledkem bylo založení mezinárodní společnosti pro výzkum neuronových sítí INNS *(International Neural Network Society)*. Univerzity po celém světě začínají zakládat nové výzkumné ústavy s tímto zaměřením.

Mnoho paradigmat neuronových sítí naráží na problém zvaný „problém proměnné stability“. Jedná se vlastně o fakt, že síť není schopna se novou informaci naučit bez poškození již dříve uložené informace. Matematik a neurobiolog dr. S. Grossberg rozpracoval síť, která má tu vlastnost, že dovede řešit „problém proměnné stability“. Adaptivní rezonanční ART sítě *(The Adaptive Resonance Theory Networks)* byly vyvinuty pro modelování mohutné paralelní architektury pro samoučící se síť k rozpoznávání obrazců. K výhodám ART sítě patří i citlivost na kontext a schopnost přiměřeně eliminovat špatné informace. [3]

Současnost

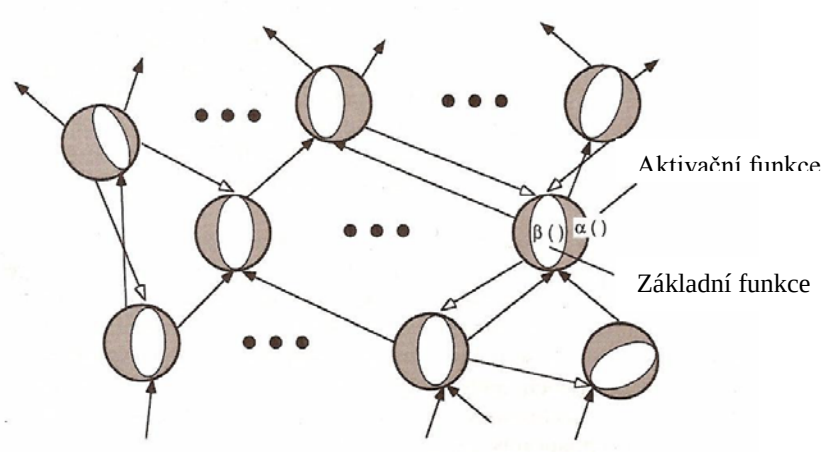
V oblasti neuronových sítí bylo dosaženo významného pokroku, což přilákalo velkou pozornost a umožnilo financovat další výzkum. Rozvoj nad rámec současných komerčních aplikací umožňuje pokračovat ve výzkumu různými směry. Technologie neuronových sítí našla uplatnění v tzv. neuročipech. [4]

2.2 CO JE NEURONOVÁ SÍŤ

Původně se termín neuronové sítě užíval ve spojitosti s biologickými neuronovými sítěmi. V technické praxi označuje matematické modely vycházející z biologických principů při zachování klíčových vlastností. Umělé sítě (ANN – Artificial Neural Networks) složené z velkého počtu úzce propojených prvků (neuronů), které se stejně jako lidé umí učit, jsou jedním z největších technologických příslibů dneška pro rozvoj informačních systémů v oblasti rozpoznávání objektů a klasifikace dat. [1]

2.2.1 Jak neuronová síť pracuje

Neuronová síť se skládá z mnoho uzlů, z nichž každý odpovídá nervové buňce ve skutečném biologickém nervovém systému, a uzlových spojení, která odpovídají nervovým propojením ve skutečném biologickém nervovém systému. Blokové schéma architektury znázorňuje obr. 2.3. Každý uzel má dvě části: jedna část (bílá elipsa) slouží k výpočtu základní funkce β() a druhá část (šedý kruh) slouží k výpočtu aktivační funkce α(). Základní funkce β() obdrží vstupní signál, který může být vstupem do sítě, nebo výstupem z jiného uzlu, a vypočítá vstupní signál pro aktivační funkci. Ta vyprodukuje výstupní signál, který může být výstupem z celé sítě nebo vstupem do jiného uzlu.



Obr. 2.3 Základní mechanismus neuronové sítě [1]

Obr. 2.3 ilustruje základní mechanismus činnosti neuronové sítě. Šipka indikuje směr toku informace (dat). Všimněme si, že šipky vycházejí z šedého kruhu a vstupují do bílé elipsy. Jsou zde dva typy šipek: černé a bílé. Předpokládejme, že signál do sítě vstupuje ze spodní části. Černé šipky znázorňují tok informací směrem vpřed a bílé šipky směrem zpět, neboli rekurzivní informace. [1]

2.2.2 Jak se neuronová síť učí

Existují dva hlavní způsoby učení neuronových sítí:

1. učení s dohledem (*supervised learning*),
2. učení bez dohledu (*unsupervised learning*).

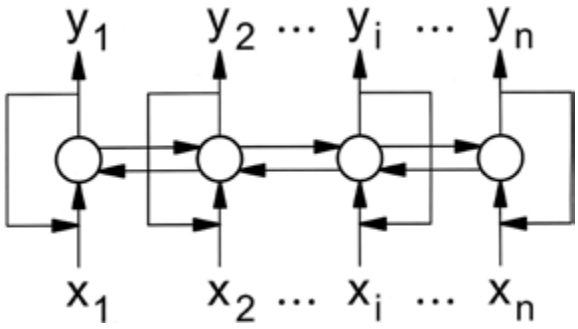
Při učení s dohledem jsou síti předkládány příklady se správnými výsledky a síť se postupně úpravou svých vah učí vracet pro každý příklad správný výsledek. Modifikací tohoto postupu je, že místo výsledku je síti sdělováno pouze to, zda se mýlí či ne.

Při učení bez dohledu jsou síti předkládána data a síť upravuje své váhy tak, aby splnila nějaké své vnitřní kritérium pro úpravu vah. Při tom dochází k takzvané samoorganizaci vah a z dat se postupně vytvářejí předem neznámé kategorie. Data určená k učení se nazývají vzory. Vzorem je většinou vektor čísel vstupujících do vstupní vrstvy sítě. Vrstva sítě je skupina neuronů vrstvy, ve kterých se mohou signály zpracovávat paralelně, tedy ve stejný okamžik. [2]

2.3 POPIS VYBRANÝCH SÍTÍ

2.3.1 Hopfieldova síť

Hopfieldova síť představuje jednovrstvou síť, ve které jsou neurony propojeny způsobem „každý s každým, kromě sebe sama“ (obr. 2.4). Hopfieldova síť obsahuje tolik neuronů, kolik je vstupů resp. výstupů neuronové sítě. Přitom každý neuron je zároveň vstupním i výstupním neuronem. Výstup každého neuronu je přes váhy spojení w_{ij} opětovně přiváděn na vstupy ostatních neuronů, čímž vzniká uzavřená smyčka (zpětná vazba). Hopfieldovu síť tedy řadíme do skupiny rekurentních (zpětnovazebních) neuronových sítí. Každý neuron přijímá jak vstupní signály, tak i interní výstupní signály od ostatních neuronů.



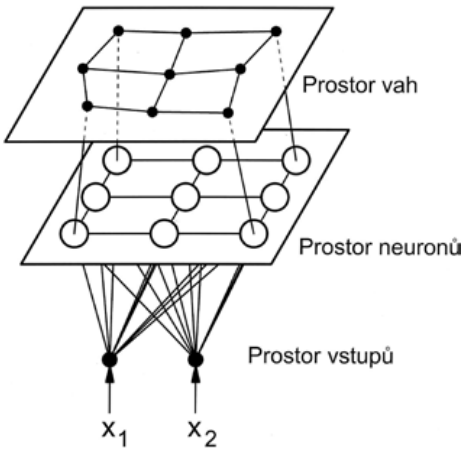
Obr. 2.4 Hopfieldova síť [5]

Během učení dochází k iterativnímu procesu. Jedná se o případ učení bez učitele, při kterém se využívá pouze reakce na předkládané vstupy. Vstupní vektor způsobí reakci na výstupech sítě a hodnoty výstupů se ihned přivádějí jako vstupy sítě. Tento proces probíhá až do stavu, kdy jsou výstupní hodnoty identické s hodnotami vstupními. Změny synaptických vah mezi jednotlivými neurony probíhají dle Hebbova algoritmu učení.

Hopfieldovi sítě dělíme na binární (s nelineární aktivační funkcí) a na spojité (např. se spojitou sigmoidální aktivační funkcí). V současné době existuje velké množství modifikací této sítě. Hopfieldova síť pracuje jako asociativní paměť, v praxi bývá používána např. k řešení optimalizačních problémů. [5]

2.3.2 Kohenova síť

Kohenova neuronová síť patří do skupiny samoorganizujících se neuronových sítí *(Self-oragnizing map)*, tzn., že ke svému učení nepotřebují učitele (jedná se o tzv. soutěžní učení). Někdy bývá nazývána též Kohonenovou topologickou mapou. Její základní schéma je uvedeno na obrázku 2.5. Kohenova síť obsahuje jedinou vrstvu neuronů v tzv. Kohenově kompetiční vrstvě. Přičemž každý vstup do sítě je plně propojen s každým neuronem v kompetiční vrstvě. Nebo-li každý neuron, nacházející se v kompetiční vrstvě má informaci o hodnotě každého vstupu. Přičemž váhy spojení každého neuronu představují souřadnice udávající konkrétní polohu neuronu v prostoru. Neurony v kompetiční vrstvě jsou mezi sebou ještě určitým způsobem laterálně spojeny. Tyto laterální spoje jsou uspořádány do předem zvolené topologické mřížky (např. čtverec, kruh atp.). Kohenova síť je tvořena formálními neurony, které nemají práh, přičemž jejich výstup je nejčastěji dvouhodnotový (0 – neaktivní; 1 – aktivní).



Obr. 2.5 Kohenova síť [5]

Základní princip učení spočívá ve stanovení „vzdáleností“ mezi předkládanými vstupními vektory a souřadnicemi neuronů umístěných v kompetiční vrstvě. Ze všech neuronů se pak vybere ten, který má nejmenší vzdálenost a stává se „vítězem“. Tento neuron se stává aktivním. Okolo vítěze se dále vytvoří okolí, do kterého se zahrnou ty neurony, které se podle

zvoleného kritéria nejvíce „podobají“ vítězi. Váhy těchto neuronů se pak modifikují. Proces učení je ukončen po vyčerpání stanoveného počtu iterací. [5]

Kohenova neuronová síť se dá matematicky vyjádřit pomocí vektoru nebo matice. Nejčastěji má struktura formu dvourozměrné čtvercové nebo obdélníkové matice, hexagonálního útvaru nebo někdy i jednorozměrného vektoru.

Matici neuronů se postupně předkládají vektory vstupního signálu (x) tak, že se zvlášť porovnává rozdíl příslušných hodnot vektoru vah (koeficientů w) každého neuronu s hodnotami vektoru vstupního signálu. K vyjádření rozdílu se může využít různých algoritmů, ale nejčastěji se dává přednost výpočtu euklidovské vzdálenosti D (2.1), tj. součet rozdílů příslušných hodnot:

$$D = (x_1 - w_1)^2 + (x_2 - w_2)^2 + ... + (x_n - w_n)^2. \qquad (2.1)$$

Výsledkem je tedy počet hodnot D, rovný počtu neuronů ve struktuře. Následně se vybere jediný neuron s nejmenším D a označí se jako tzv. vítěz (winner).

Váhy W (2.2) vítězného neuronu se pak upravují (updatují), aby se co nejvíce přiblížily hodnotám právě předloženého vstupního vektoru (x).

$$W_{i\ nové} = W_{i\ staré} + \alpha\ (x - W_{i\ staré}), \qquad (2.2)$$

kde α je učící koeficient vyjadřující rychlost učení (může nabývat hodnot 0 až 1, např. $\alpha = 0.6$), W_i je vektor vah (koeficientů) i-tého neuronu $W_i = [W_{i1}, W_{i2}, ..., W_{in}]$ a x je vstupní učící vektor $x = [x_1, x_2, ... x_n]$.

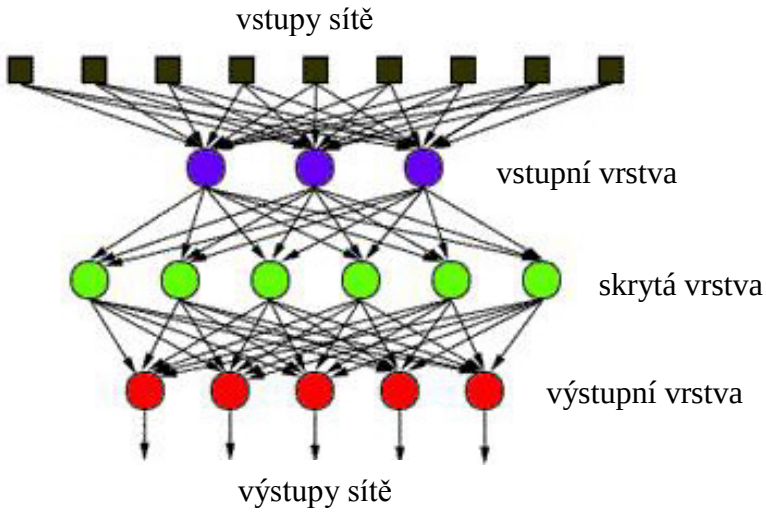
Nevýhodou, která platí i obecně pro většinu neuronových sítí, je vysoká výpočetní náročnost, kdy uvedený výsledek byl naučen až po 10 000 předložených vstupních vektorech. Rychlost lze sice zvýšit volbou vyšších hodnot koeficientů α, příp. β, více blížící se hodnotě 1, ale to pak může snížit přesnost a kvalitu učení. Je tedy nutné nalézt kompromis.

Kohenovy mapy jsou silným a kvalitním nástrojem pro identifikaci neznámých vlastností a parametrů, skrytých v digitalizovaných vzorcích libovolného signálu. V některých aplikacích mohou pracovat jako alternativa

k jiným algoritmům, v některých aplikacích jsou již nenahraditelné. S použitím programu Matlab a jeho toolboxu Neural Networks, lze vytvořit jednoduchý rozpoznávač samohlásek v nahraném řečovém signálu. [9]

2.3.3 MLP síť

MLP *(Multi-Layer-Perceptron)* je vícevrstvá neuronová síť. Skládá se obvykle ze tří vrstev: vstupní vrstvy, alespoň jedné skryté (prostřední) vrstvy a výstupní vrstvy (obr. 2.7). Na rozdíl od IAC a Hopfieldových sítí jsou váhy v MLP jednosměrné a vedou ze vstupní do skryté vrstvy a ze skryté do výstupní vrstvy. Neurony v sousedních vrstvách jsou většinou propojeny každý s každým. Výstup MLP můžeme označit jako klasifikační rozhodnutí. Na rozdíl od sítě IAC, kam se musí informace zakódovat natvrdo pomocí námi stanovených vah, MLP své váhy přizpůsobuje automaticky pomocí učícího algoritmu datům, které mu předložíme k naučení.



Obr. 2.6 Architektura MLP sítě [8]

Trénování sítě probíhá tak, že vytvoříme soubor trénovacích dat, a síť tento soubor několikrát po sobě zpracovává. Trénovací soubor se skládá z trénovacích vzorů, kde každý vzor představuje vektor vstupních hodnot a

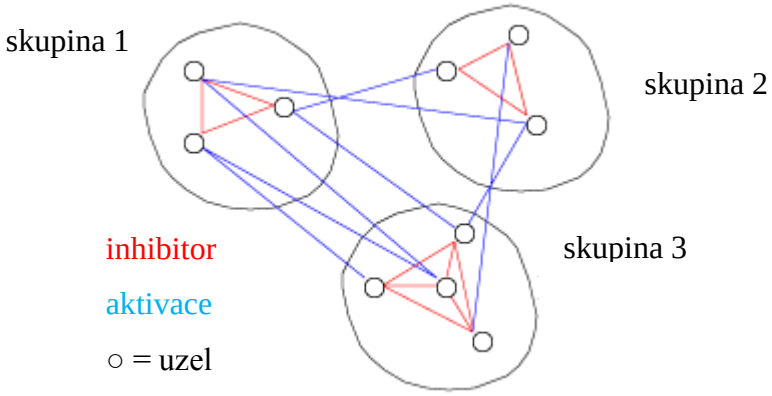
vektor požadovaných výstupních hodnot. Učící algoritmus mění koeficienty sítě tak, aby se odchylka výstupů sítě od žádaných výstupů při předložení příslušných vstupních hodnot minimalizovala. Nejprve je po předložení vstupního vzoru postupně po vrstvách vypočten výstup hodnot všech neuronů pro každou nevstupní vrstvu. Tento výstup je porovnán se žádaným výstupním vzorem načteným z trénovacího souboru. Informace o odchylce je sítí šířena směrem od výstupní vrstvy k vrstvě vstupní a jsou podle ní upravovány váhy sítě tak, aby až síť bude stejnou dvojici vzorů číst příště, byla odchylka na její výstupní vrstvě menší.

Poté, co se síť správně naučí trénovací soubor, může být testována na jiném souboru dvojic vstupních a výstupních vektorů, abychom zjistili, jak dobře se síť naučila zobecňovat. Při trénování sítě se sleduje i její výkon. Trénink je ukončen při dosažení určitého počtu etap nebo poklesne-li chyba výstupu sítě pro tréninkovou množinu pod žádanou hranici. [2]

2.3.4 Síť pro interaktivní aktivaci a soutěžení

Síť IAC *(The Interactive Activation and Competition Network)* se skládá ze skupin vzájemně soutěžících jednotek (neuronů), kde každá jednotka představuje nějakou mikrohypotézu nebo vlastnost (obr. 2.8). Jednotky ve společné skupině představují vzájemně se vylučující vlastnosti a váhy, kterými jsou mezi sebou propojeny a jsou proto záporné. Pro spoje mezi jednotkami patřícími do různých skupin platí, že jsou na nich kladné váhy, pokud vlastnosti představované jimi propojenými jednotkami se vzájemně nevylučují. Všechny spoje jsou obousměrné.

Hodnota patřící určité jednotce se nazývá její aktivace. Jednotka s vysokou aktivací je aktivní. Vytvořenou síť lze použít k vyvolávání do ní uložených informací. Aktivace každé jednotky může být považována za míru



Obr. 2.7 IAC diagram [6]

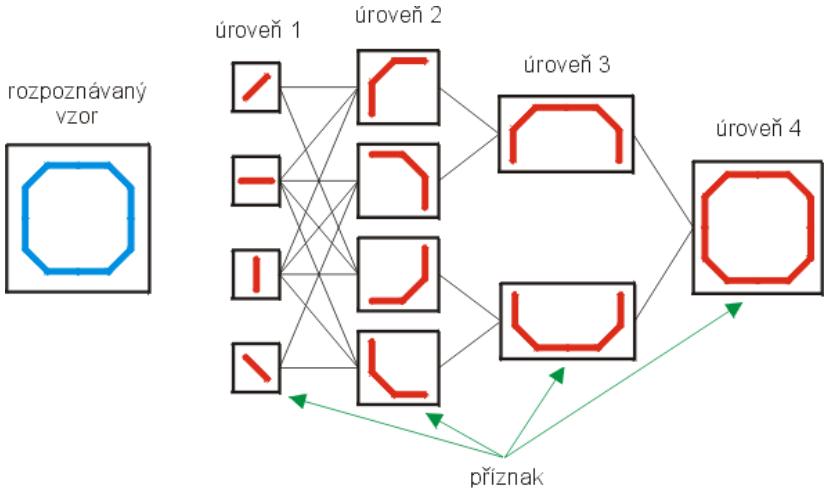
důvěry v hypotézu nebo vlastnost, kterou tato jednotka představuje. Váhy mezi jednotkami v síti indikují, jak silně víra v jednu hypotézu implikuje víru v jinou hypotézu. Mechanismus IAC je cyklický proces aktualizace víry v hypotézu v závislosti na aktuálních důkazech.

V síti, která cyklicky upravuje aktivace svých jednotek podle těchto pravidel, lze pozorovat dva hlavní jevy. Jednotky napojené klanými vahami na nejaktivnější jednotku zvyšují svoji aktivaci a jednotky napojené zápornými vahami na nejaktivnější jednotku své aktivace snižují. [2]

2.3.5 Neocognitron

Neocognitron je vícevrstvá hierarchická neuronová síť vytvořená pro řešení úloh při rozpoznávání rukou psaných znaků. Hlavní výhodou této sítě je schopnost rozpoznávat nejen naučené vzory, ale i vzory, které z nich vzniknou částečným posunutím, otočením nebo i částečnou deformací.

Základním principem fungování neuronové sítě neocognitron je hierarchická detekce příznaků (obr. 2.9), která spočívá v rozdělení detekovaných příznaků do několika úrovní. V první úrovni se detekují nejjednodušší příznaky a v každé následující úrovni jsou detekované příznaky komplexnější. Pro detekci příznaků v určité úrovni se využívají informace získané v předcházející úrovni. Celkový počet úrovní závisí na složitosti rozpoznávaného vzoru. [7]



Obr. 2.8 Neocognitron - hierarchická detekce příznaků [7]

2.4 ZHODNOCENÍ MODELŮ PRO ÚČELY ROZPOZNÁNÍ ŘEČI

Hopfieldovy sítě se používají především k rekonstrukci neúplných a šumem poškozených dat a k optimalizaci systémů. Jejich nevýhodou jsou veliké nároky na paměť, což může způsobit chybu při identifikaci předloženého vzoru.

Kohenova síť je vhodná k analýze dat a k vytváření sémantických map. Nevýhodou je vysoká výpočetní náročnost a schopnost jejího učení pouze v mládí.

Neocognitron je neuronová síť vytvořená za speciálním účelem rozpoznávání rukou psaných znaků.

Vícevrstvá neuronová síť typu MLP je vhodná pro řešení fonetické transkripce, klasifikaci obrazů, hodí se k aproximaci funkcí a predikaci časových řad.

3. ROZPOZNÁVAČE PŘIROZENÉ ŘEČI

Jedním z možných využití rozpoznávačů řeči je dekódování akustických řečových signálů na jazykové prvky, jako jsou slova a věty a také převádění mluvené řeči na strojovou. Přirozeně je snaha o co nejpřesnější rozpoznání řečových pojmů a porozumění mluvené řeči. Zjištěné poznatky se dále využívají k vytvoření matematického modelu rozpoznávače nebo vlastního rozpoznávače. Mechanismům zpracovávajícím mluvenou řeč je věnována velká pozornost, přesto její analýza není stále dokonalá. I tak je už součástí našeho denního života.

Původně se používaly k rozpoznávání řeči jiné systémové struktury než umělé neuronové sítě, jako například distanční klasifikátor používající modely a pravděpodobnostní klasifikátor založený na Markovových modelech (HMMs).

Strojové rozpoznávání řeči, běžně nazývané jako přepis mluvených slov nebo vět, je pro vědce velkou výzvou. Je založeno na teorii rozpoznávání vzorů, stejně jako celá kategorie aplikací zahrnující rozpoznání řeči mluvčího i problémy týkající se syntézy řeči, rozpoznání obrazu, časového sledu a jednotlivých znaků. Tato kategorie problémů je často řešena za použití jednoduchého, ale účinného principu učení s dohledem. [1]

Úspěšnost rozpoznávání z velké části závisí na podmínkách, v nichž je rozpoznávač testován. Obecně platí, že čím silnější omezení jsou kladena na formu rozpoznávané řeči, tím je rozpoznávání pro počítač snazší a tudíž úspěšnější. V závislosti na těchto omezeních je možné rozpoznávače kategorizovat podle následujících kritérií:

Velikost slovníku

Čím více slov může obsahovat rozpoznávaná promluva, tím je proces rozpoznávání výpočetně náročnější a roste pravděpodobnost záměny rozpoznávaných slov.

Jeden nebo více řečníků

Systémy nastavené na rozpoznávání promluv jedné konkrétní osoby dosahují větší úspěšnosti rozpoznávání než systémy rozpoznávající promluvy libovolného řečníka.

Izolovaná slova vs. plynulá řeč

Rozpoznávání plynulé řeči, kde nejsou snadno zjistitelné hranice slov, je výrazně náročnější než rozpoznávání izolovaných slov (např. povelů).

Doménová a jazyková omezení

Doménová omezení se vztahují k úkolu, ke kterému má být rozpoznávač použit. Omezení mohou být sémantická nebo syntaktická, která se snadněji aplikují na jazyky s pevným pořádkem slov.

Čtená vs. spontánní řeč

Spontánní řeč je pro zpracování strojem obtížnější, protože může obsahovat nečekané pauzy, přeřeknutí a nekompletní věty.

Nepříznivé podmínky

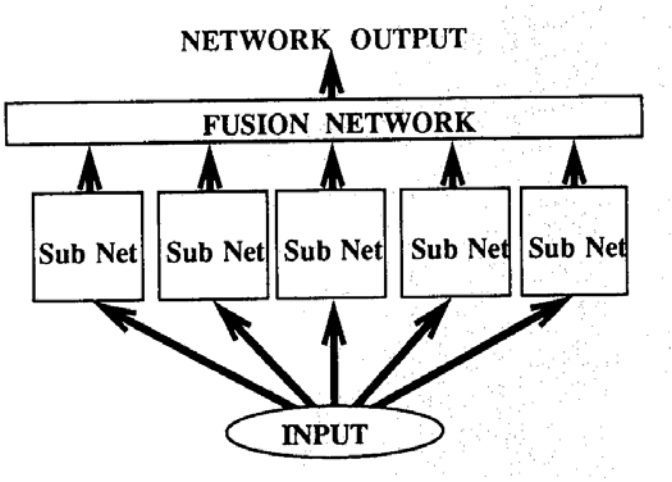
Rozpoznání může být výrazně ztíženo šumem, akustickými podmínkami prostoru, elektrickými vlastnostmi kanálu, použitím různých mikrofonů, frekvenčním omezením, nebo změnou stylu mluvy. [10]

Následující kapitoly popisují několik vybraných rozpoznávačů přirozené řeči.

3.1 NETWORK FUSION

Neuronové sítě s dopravním zpožděním, které se používají pro rozpoznávání řeči, jsou navrženy tak, aby byly neuronové váhy modifikovány učícím se algoritmem, který minimalizuje míru výkonu nad množinou zkušebních dat. Typický proces návrhu neuronové sítě vyžaduje specifikaci síťové architektury a její naučení pomocí všech dostupných dat. To může trvat velmi dlouho a výkon takové sítě pro řešení komplexních problémů, jako je rozpoznání řeči, je mnohdy nedostačující díky svojí neschopnosti daný problém zobecnit. Další nevýhodou tohoto návrhu je, že znemožňuje zahrnutí jakékoliv apriorní znalosti o problému během učení.

Mnohem vhodnější je navrhovat velké neuronové sítě tak, že je rozdělíme na několik menších, z nichž každá bude zaměřena na řešení dílčího problému. Tyto podsítě pracují s menším objemem dat, rychleji se učí a lépe umožňují zobecnění problému. Velké neuronové sítě tedy vznikají fůzí několika menších, nezávislých, již naučených podsítí. Pro tyto sítě je užíván termín *network fusion* (obr. 3.1).



Obr. 3.1 Proces vytvoření sítě technikou *network fusion* [11]

3.1.1 Experiment s network fusion

Databáze použitá pro tento experiment zaměřený na rozpoznávání řeči obsahovala tři vstupní slova ve španělštině: uno, dos, tres. Byla zaznamenána běžnými telefonními linkami ve španělském Madridu. Na tvorbě databáze se podílelo více než 1000 mluvčích, kteří nebyli seznámeni s technologií rozpoznávání hlasu. Odstup signálu od šumu nahrávané řeči se pohyboval v rozmezí od -17.0 dB do 46.0 dB. Řeč byla limitována pásmem od 300 Hz do 3.4 kHz se vzorkovací periodou 8 kHz. Databázi tvořilo 3044 vzorků řeči, z nichž 1685 bylo vytvořeno muži a 1359 ženami. Databáze byla rozdělena na 1850 vzorků k učení sítě a 1194 vzorků pro její otestování.

Cílem bylo vystavět síť s dopravním zpožděním se 420 ms plovoucím oknem schopnou rozpoznávat slova z vytvořené databáze. Jestliže se v plovoucím okně objevilo slovo shodné se slovem v databázi, výstupem sítě byl kladný impuls, v opačném případě záporný. Slovo bylo správně rozpoznáno tehdy, když hodnota maximálního kladného impulsu na výstupu neuronové sítě se shodovala s hodnotou impulsu odpovídajícího obrazu řeči v okně.

Každé vstupní slovo bylo zkoumáno ve třech akustických polohách: *onset* (hrubá), *center* (podrobnější), *end* (určující). Výsledek zkoumání je shrnut do tabulky 3.1.

Výkon podsítě			
Síť	**Chyby**	**Nerozhodnuto**	**Pravděpodobnost chyby**
onset	267	30	0.236
center	120	42	0.107
end	289	16	0.252
Výkon sloučených podsítí			
onsetF	179	138	0.175
centerF	99	44	0.089
endF	164	98	0.155

Tabulka 3.1 Výkony podsítí [11]

Sloupec „Chyby“ udává počet nezjištěných chyb, zatímco sloupec „Nerozhodnuto“ udává počet slov, o nichž síť nerozhodla. Horní část tabulky se týká separovaných podsítí a ukazuje velkou chybovost při rozpoznávání slov. Její dolní část představuje sníženou chybovost při spojení podsítí ve větší celky.

Tabulka 3.2 dokumentuje, že absolutně nejmenší chybovost vykazují sítě vzniklé spojením co největšího počtu podsítí metodou *network fusion*. [11]

Výkon dvouvrstvé sítě			
Databáze	**Chyby**	**Nerozhodnuto**	**Pravděpodobnost chyby**
Trénovací	31	92	0.029
Testovací	16	48	0.027

Tabulka 3.2 Výkon finální sítě [11]

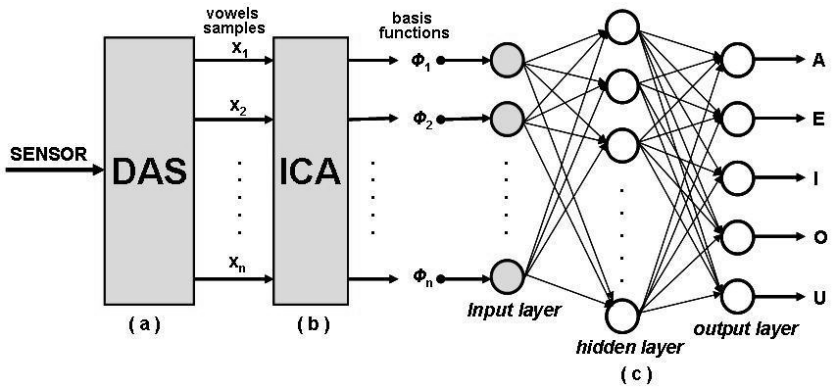
3.2 SUBVOKÁLNÍ KOMUNIKACE

Nevýhodou běžných systémů pro rozpoznávání řeči je jejich nezpůsobilost rozpoznat subvokální komunikaci. Jejich výkon může být silně ovlivněn i mnoha faktory jako jsou úroveň šumu okolního prostředí a odraz. Tyto systémy jsou samozřejmě založené na rozpoznávání slyšitelné řeči, jejíž hlasitost může být v některých případech nedostačující.

K řešení těchto problémů byly vyvinuty systémy pro rozpoznávání řeči založené na zaznamenávání subvokální komunikace. Například, aby bylo možné řídit softwarové rozhraní, Jorgensen a Binsted použili pro získání subvokálů elektromyogram (EMG), z něhož vlnovou analýzou extrahovali charakteristiky a s použitím neuronové sítě subvokály rozpoznali. Nevýhodou tohoto řešení je použití fixních vlnových filtrů, které nevyužívají statistickou informaci ze vstupních dat, takže neobsahují důležité informace pro rozpoznávání.

Tento nedostatek odstraňuje nezávislá dílčí analýza (ICA) založená na statistikách vyšších řádů (HOA), v důsledku čehož extrahované charakteristiky dobře reprezentují vstupní data.

Blokový diagram systému využívajícího analýzu ICA je znázorněn na obrázku 3.2.



Obr. 3.2 Návrh systému subvokálního rozpoznávání řeči [12]

Systém pracuje ve třech fázích: získávání subvokálů, fáze učení a klasifikace. První fáze využívá *Data Acquisition System* (DAS), fáze učení probíhá pomocí ICA analýzy a klasifikační fáze využívá neuronových sítí.

Sensitivita *(Sens)* (3.1) a specifičnost *(Spec)* (3.2) jsou nejrozšířeněji používané statistiky při vyhodnocování diagnostických testů

$Sens = TP/(TP+FN)$, kde (3.1)

TP…true-positive diagnosis

FN…false-negative diagnosis

$Spec = TN/(TN+FP)$, kde (3.2)

TN…true-negative diagnosis

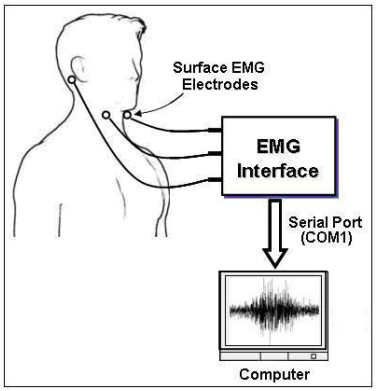
FP…false-positive diagnosis

Přesnost *(Accu)* (3.3) vyjadřuje vztah mezi měřenou hodnotou a standardem a byla stanovena vztahem

$$Accu=(TN+TP)/(TP+FP+FN+TN \quad (3.3)$$

3.2.1 Experiment se subvokální komunikací

V experimentu se pracovalo s databází samohlásek získaných z portugalštiny a vyjádřených pomocí EMG signálů (obr. 3.3). Databáze obsahovala 150 vzorků každé fonémické hlásky od deseti mluvčích (7 mužů, 3 ženy) ve věkovém rozpětí od 18 do 48 let.



Obr. 3.3 Schéma subvokálního systému [12]

Následující tabulka 3.3 ukazuje úspěšnost metody fonémického rozpoznávání v procentech.

Fonémická hláska	**A(/a/)**	**E(/ɛ/)**	**I(/i/)**	**O(/ɔ/)**	**U(/u/)**
A(/a/)	35	3	0	0	0
E(/ɛ/)	0	34	4	0	0
I(/i/)	0	0	35	3	0
O(/ɔ/)	2	0	0	34	1
U(/u/)	0	0	0	0	36
Úspěšnost (%)	94,6%	91,8%	94,6%	91,8%	97,3%

Tabulka 3.3 Výkon finální sítě [12]

Výpočtem zjistíme, že průměrná úspěšnost této metody je 93,99%, přičemž specifičnost dosahovala 93,92% a sensibilita 94,05%.

Dříve navržené sítě	Použitá metoda	Úspěšnost
Rozeznávání malých slov s použitím povrchové elektromyografie v akusticky nepříznivém prostředí	- HMM - Neuronová síť	74%
Rozeznávání podprahové řeči založené na EMG/EPG signálech	- HMM - Neuronová síť	92%
Rozeznávání podprahové řeči	- HMM -Wavelet	92%
Práce s webovým prohlížečem s použitím EMG založeném na subvokálním rozpoznávání řeči	- HMM -Wavelet	92%

Tabulka 3.4 Dříve navržené sítě [12]

Srovnáním výsledků popsaného experimentu s hodnotami uvedenými v tabulce 3.4 zjistíme, že metoda subvokálního rozpoznávání řeči založená na EMG signálu s použitím ICA a MLP neuronových sítí je úspěšnější než metody dříve používané. [12]

3.3 GENEROVÁNÍ CÍLŮ DOPŘEDNÝM/ZPĚTNÝM ŠÍŘENÍM PRAVDĚPODOBNOSTI

Hybridní přístupy k rozpoznávání řeči jsou motivovány zkoumáním skrytých Markovových modelů *(Hidden Markov Models)* a neuronových sítí *(Neuron Networks)*. [13] Skrytý Markovův model generuje v krátkých časových okamžicích náhodnou posloupnost pozorování (posloupnost vektorů pozorování). V každém časovém kroku změní model svůj stav, a to podle souboru předem daných pravděpodobností přechodu. Stav, do kterého model přejde, vygeneruje jedno pozorování (jeden vektor pozorování) podle rozdělení výstupní pravděpodobnosti příslušné k tomuto stavu.

Silná stránka použití skrytých Markovových modelů spočívá v existenci efektivních algoritmů pro estimaci jejich parametrů a algoritmů používaných při rozpoznávání řeči [14], zatímco neuronové sítě se ukázaly být mocným nástrojem pro statistickou klasifikaci úloh. Cíle pro trénování neuronových sítí určených k rozpoznávání řeči jsou estimovány novými metodami pomocí obecného *forward-backward* algoritmu.

Testy porovnávající běžně používané systémy založené na Viterbiho algoritmech a hybridními systémy, vycházejí jednoznačně lépe pro hybridní systémy. Chybovost se u hybridních systémů snížila až o 15%.

Jako u mnoha ostatních hybridních systémů, tak i zde jsou neuronové sítě použity jako generátor posloupnosti pravděpodobnostních stavů. V této studii byl použit třívrstvý model neuronové sítě. Jako modelové jednotky byly použity fonémy, každý foném má od jednoho do tři stavů, kde každý stav odpovídá jednomu výstupnímu uzlu neuronové sítě.

Na rozdíl od mnoha již existujících hybridních systémů, které nemodelují vnitřní přechodné modely fonémů, tento nový hybridní systém využívá k rozpoznávání řeči dvojnásobný pravděpodobnostní proces.

3.3.1 Experiment s dopředným/zpětným šířením pravděpodobnosti

Experiment je založen na číselných sekvencích, které byly získány z veřejné telefonní sítě. Každá promluva obsahuje 1 až 6 plynule vyslovovaných číselných řetězců. V databázi se objevují chybné začátky, pauzy, opakování a běžné zvuky z okolí. Seznam slov obsahuje: *zero, oh, one, two, three, four, five, six, seven, eight, nine.* Data (slova) byla náhodně rozdělena do tří skupin: tréninkový set se 2090 slovy, testovací set s 500 slovy a vyhodnocovací set s 1600 slovy.

3.3.1.1 HMM systém

Pro rozpoznávání fonémů z databáze byly použity srovnávací Markovovy modely HTK *(Hidden Markov Model Toolkit).* Každý foném byl reprezentován třístavovým *left-to-right* modelem složeným ze čtyř parametrového Gaussova směsového modelu *(Gaussian Mixture Model)* využívajícího diagonální vzájemnou vazbu. Řečový signál byl parametrizován každých 12,8 ms Hammingovým oknem s parametrem 25,6. Stavový vektor je 26-dimensionální tvořený 12 LPC-cepstrálními koeficienty, normalizovanou energií a rozdíly mezi nimi (použita metoda CMS *(Cepstral Mean Subtraction)*). Celkový počet kontextově závislých fonémových modelů byl 77.

3.3.1.2 Hybridní systémy

Pro hodnocení metod generujících cíle byly vyvinuty dva hybridní systémy – The Baseline, The New System. První z nich je založený pouze na 0-1 cílech a ke každému vstupnímu řečovému vektoru existuje pouze jedna třída s nenulovým cílem. Druhý systém byl přeučen pomocí *forward-backward* cílů. Rozdíl mezi oběma systémy je v tom, že druhý navíc využívá *forward-backward* algoritmu pro výpočet pravděpodobnosti změny vnitřních fonémů.

3.3.2 Vyhodnocení experimentu

Výsledky experimentu se systémy HMM, Baseline a New System jsou shrnuty v tabulce 3.5. Sledovali jsme dopad modelování vnitřních modelů přechodných stavů na výkon systémů. Výsledkem je zjištění, že modelování má malý vliv na správné rozpoznání jednotlivých slov, ale snižuje chybovost o 30%, nicméně přesnost porozumění větám se o 5% zvýšila. [13]

Datový set	Jednotka	HMM	Baseline	New
Vytvořený set	Slovo	4.1%	4.1%	3.1%
	Slovní spojení	13.2%	13.0%	11.4%
Testovací set	Slovo	5.7%	6.0%	4.9%
	Slovní spojení	18.9%	19.7%	16.7%

Tabulka 3.5 Porovnání chybovosti systémů [13]

4. ANALÝZA ŘEČOVÉHO SIGNÁLU

Analýzou řečového signálu rozumíme proces analýzy a rozpoznávání elementů řečového signálu, při němž počítač (robot) převádí vstupní akustický signál do textové podoby.

4.1 CHARAKTERISTIKA LIDSKÉ ŘEČI

Lidskou řeč lze charakterizovat pomocí:

- akustické struktury – amplitudově-frekvenční spektrum měnící se v čase
- lingvistické struktury – gramatika a skladba
- subjektivního vlivu osobnosti řečníka – intonace, rytmus, barva hlasu

Jednotlivé zvuky lidské řeči, ze kterých se skládají slova daného jazyka, jsou hlásky. Vznikají tím, že vydechovaný proud vzduchu různým způsobem v hrdle a v ústech upravujeme. Frekvencí kmitů hlasivek je charakterizován základní tón lidského hlasu, který se u většiny lidí pohybuje v rozmezí 150 – 400 Hz. Hlásky dělíme na samohlásky a souhlásky. [15]

4.1.1 Samohlásky

V akustickém spektru každé samohlásky se objevují zesílené tóny, které vznikají rezonancí v dutině hlasového traktu. Tyto zesílené tóny nazýváme *formanty*. V češtině se můžeme setkat s pěti samohláskami, které jsou krátké i dlouhé: a, e, i, o, u – á, é, í, ó, ú.

4.1.2 Souhlásky

Souhlásky jsou takové hlásky, jejichž charakteristickým rysem (na rozdíl od samohlásek) je šum, který vzniká specifickým postavením či pohybem mluvidel. Souhlásky se dělí podle účasti hlasivek na výslovnosti na znělé, při nichž se hlasivky chvějí, a na neznělé, při nichž se hlasivky nechvějí.

- znělé: b, v, d, t, ď, z, ž, g
- neznělé: p, f, t, ť, s, š, k

Souhlásky lze dále dělit i podle způsobu výslovnosti a podle místa výslovnosti.

Střídáním poloh hlasového ústrojí podle předem definovaného pořadí vznikají těmto polohám odpovídající akustické signály, ze kterých se formují zvukové elementy *fonémy* umožňující rozlišovat zvukovou podobu slova. Zjednodušeně lze říci, že pro češtinu odpovídají jednotlivé fonémy vysloveným hláskám.

Problém spočívá v tom, že jednotlivé fonémy se mění vyslovením v různém kontextu. Jejich akustická realizace závisí jak na předcházejícím a následujícím zvuku, tak i na tempu a intonaci řeči. Tato závislost, zvaná jako *koartikulace*, značně komplikuje složitost dalších postupů zpracování, ať už pro účely hlasové syntézy, nebo rozpoznávání řeči. [16]

4.2 SYSTÉMY NA ROZPOZNÁVÁNÍ ŘEČI

Podle závislosti na řečníkovi rozeznáváme tyto typy systémů pro analýzu řeči:

závislé na řečníkovi - jsou vyvíjeny pro práci s jedním řečníkem. Jsou to nejjednodušší typy systémů, jejich vývoj je poměrně snadný, jsou levnější a většinou přesnější, ale nejsou tak flexibilní jako další typy.

nezávislé na řečníkovi - jsou schopny pracovat s jakýmkoli řečníkem určitého typu (např. řečníkem mluvícím česky, anglicky, apod.). Vývoj těchto systémů je nejsložitější, systémy jsou nejdražší a nejsou tak přesné jako systémy závislé na řečníkovi. Nicméně jsou flexibilnější a mají větší rozsah použití, než systémy závislé na řečníkovi.

adaptivní - jsou vyvíjeny s cílem přizpůsobovat se vlastnostem nových řečníků. Jejich složitost je někde mezi systémy nezávislými a systémy závislými na řečníkovi.

4.3 POSTUP PŘI ROZPOZNÁVÁNÍ ŘEČI

a) **vzorkování** – provádí se „klasickými metodami“ (mikrofon, zesilovač, D/A převodník).
b) **zpracování akustického signálu** – techniky spektrální analýzy (LPC, MFCC, cochlea modelling, …).
c) rozpoznání fonémů, skupin fonémů a slov. Principem je přiřazení vstupu nějakému slovu ze slovníku známých slov (výstup). Techniky pro rozpoznání: DTW, HMM, ES, NS, kombinace těchto technik.
d) Některé systémy se snaží *porozumět* řeči, tzn., snaží se zkonvertovat slova do reprezentace, která umožní zjistit, co nejpřesnější interpretaci řečníkova sdělení. Jinými slovy, provádějí syntaktickou a sémantickou analýzu. [16]

5. VLASTNÍ NAVRŽENÝ SYSTÉM

Pro realizaci bylo zvoleno programovací prostředí MATLAB a dopředná dvouvrstvá neuronová síť typu SOM *(Self-Organizing Map).*

5.1 ARCHITEKTURA

Architektura systému vychází z neuronové sítě typu Kohonenovy samoorganizační mapy, která je architekturou vycházející ze strategie soutěžního učení, tzn. učení bez učitele. Provádí druh shlukové analýzy a používá se pro klasifikaci velké skupiny dokumentů, v našem případě skupiny hlásek.

5.2 ANALYZOVANÁ DATA

Pro účely trénování neuronové sítě byly vytvořeny vzory hlásek v digitalizovaném formátu wav. Vzory hlásek byly nahrány pomocí mikrofonu AKG D5 s použitím software Sony Sound Forge 9.0. Každý vzorek je nahrán s vzorkovací frekvencí 44 100 Hz a bitovým rozlišením 16 bitů / vzorek.

Testovací sada hlásek obsahuje:

- samohlásky: a, e, i, o, u, y
- sykavky: c, f, s, x, z
- souhlásky: b, d, g, h, ch, j, k, l, m, n, p, r, t, v

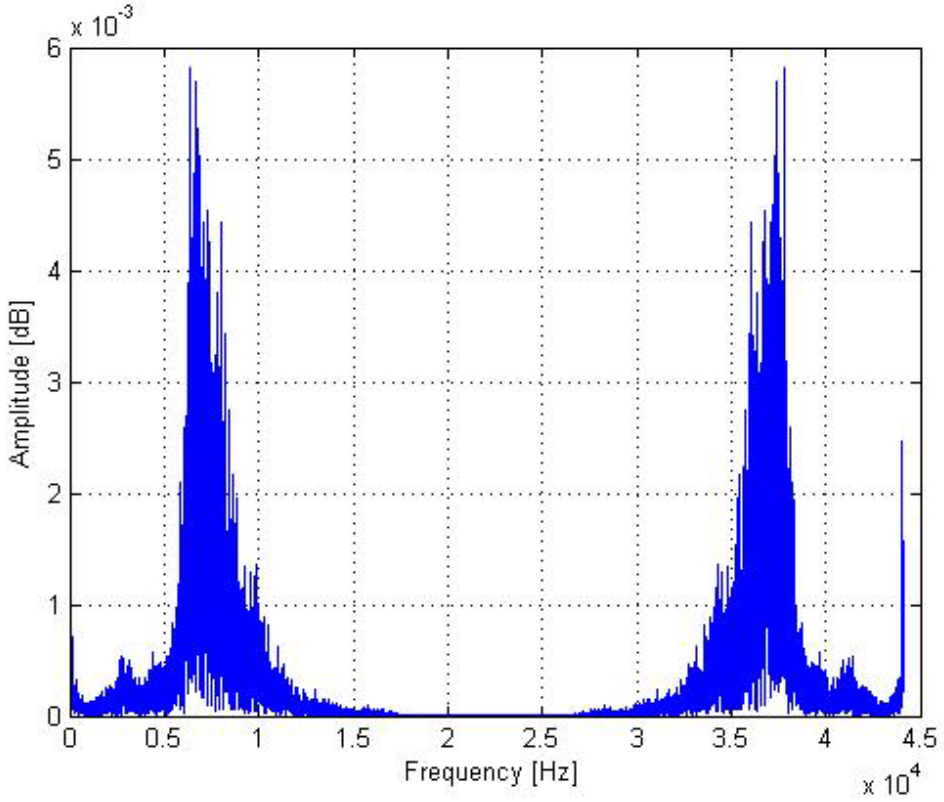
5.2.1 Zpracování získaných dat

Získaná data (vzorky hlásek) byla pomocí funkce *wavread()* importována do programovacího prostředí MATLAB. Pomocí algoritmu FFT *(Fast Fourier Transform)* byla poté provedena spektrální analýza jednotlivých datových vzorků.

Algoritmus FFT je definován funkcí:

$$X(k) = \sum_{j=1}^{N} x(j)\omega_N^{(j-1)(k-1)} \quad \text{,kde} \quad \omega_N = e^{(-2\pi i)/N} \qquad (5.1)$$

Po průchodu vzorků Fourierovou transformací byla získána jejich frekvenční spektra. Jak je patrno z obr. 5.1, výsledné frekvenční spektrum každého vzorku lze rozdělit na dvě stejné části, jež jsou zrcadlově obrácené. Z toho důvodu jsem použil pro klasifikaci vzorků pouze jednu polovinu jejich frekvenčního spektra *(One-sided Spectrum)*. Jako hraniční frekvenci jsem určil frekvenci f=15 000 Hz.



Obr. 5.1 Zrcadlové spektrum signálu hlásku „s“

Dalším krokem realizace bylo převedení frekvenčního spektra signálu do takového formátu, který by již byl vstupem do neuronové sítě. Signál byl tedy rozdělen na deset stejných dílů a z každého z nich vypočtena střední

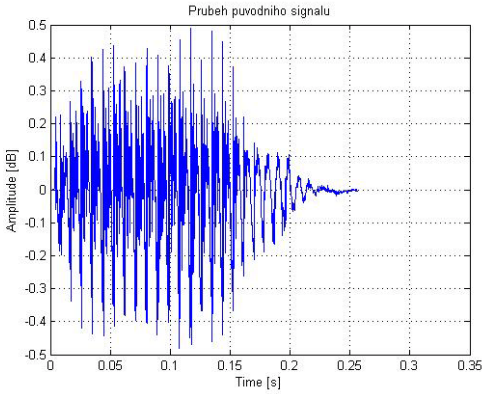
hodnota signálu. Tím bylo získáno deset číselných hodnot, které reprezentují příslušný vzorek a jsou vstupem do neuronové sítě.

5.2.2 Výsledná frekvenční spektra

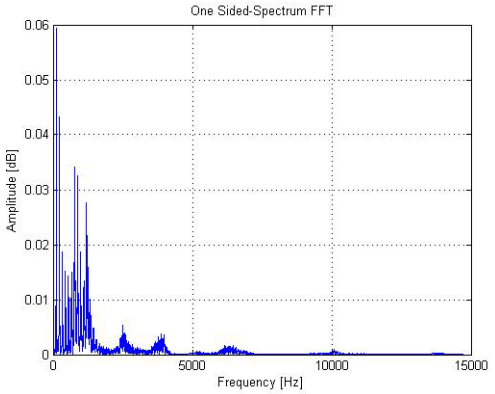
Frekvenční spektra jednotlivých vzorků se liší podle toho, zda jde o souhlásky znělé, neznělé, tvrdé, měkké nebo sykavky. Zmíněné rozdíly uvádím na jednotlivých příkladech.

5.2.2.1 Samohlásky

Samohláska „a“:

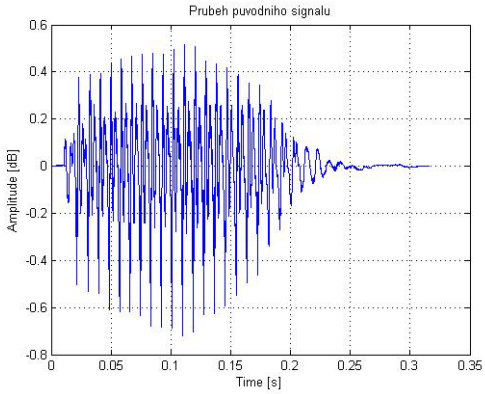


Obr. 5.2 Původní signál samohlásky „a“

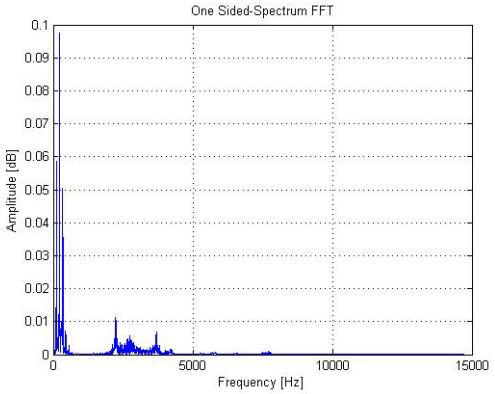


Obr. 5.3 Signál samohlásky „a“ po FFT

Samohláska „i“:



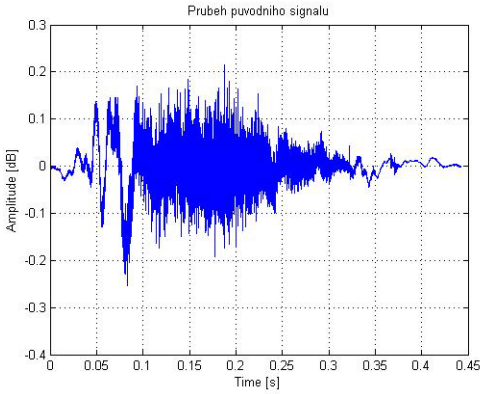
Obr. 5.4 Původní signál samohlásky „i“



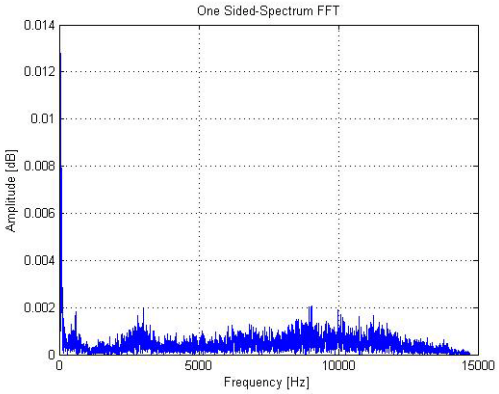
Obr. 5.5 Signál samohlásky „i“ po FFT

5.2.2.2 Sykavky

Sykavka „f“:

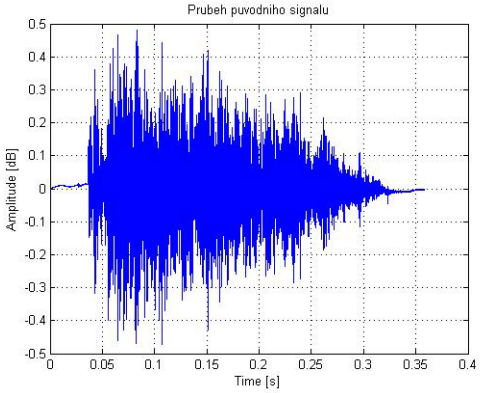


Obr. 5.6 Původní signál sykavky „f“

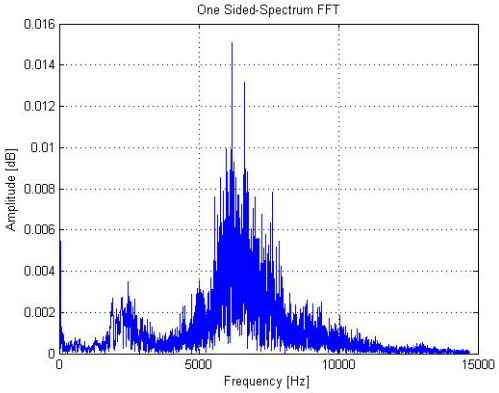


Obr. 5.7 Signál sykavky „f“ po FFT

Sykavka „x“:



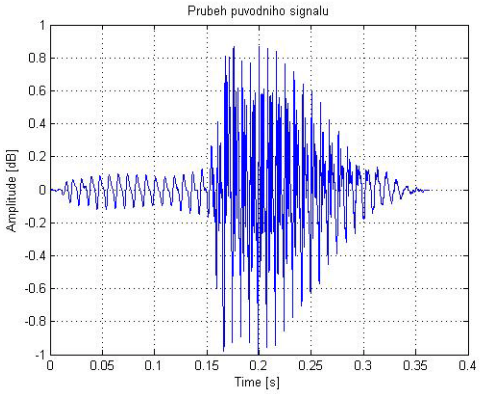
Obr. 5.8 Původní signál sykavky „x“

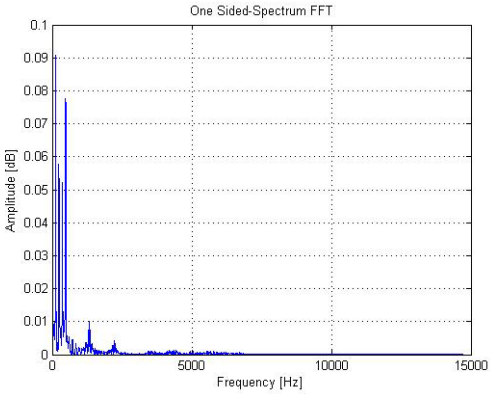


Obr. 5.9 Signál sykavky „x“ po FFT

5.2.2.3 Souhlásky

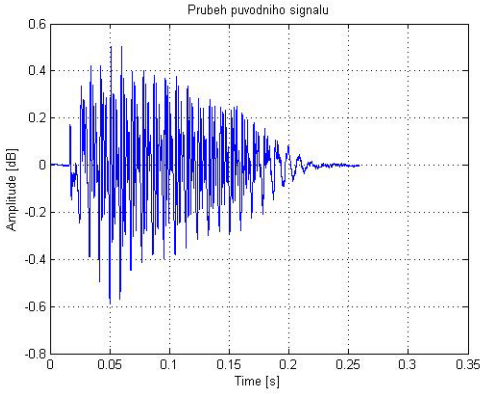
Souhláska „b“:

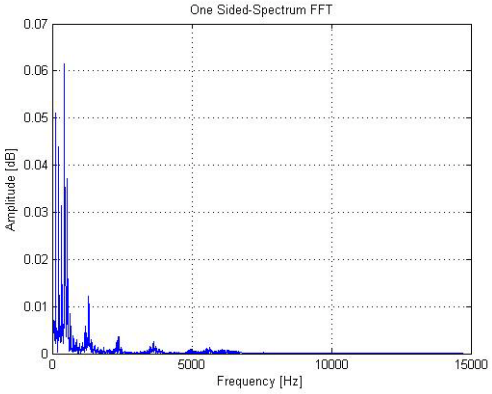




Obr. 5.10 Původní signál souhlásky „b“ Obr. 5.11 Signál souhlásky „b“ po FFT

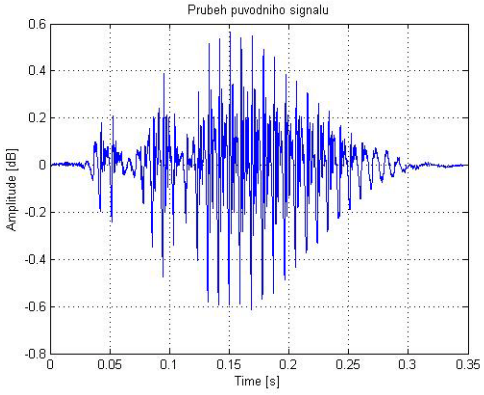
Souhláska „p“:

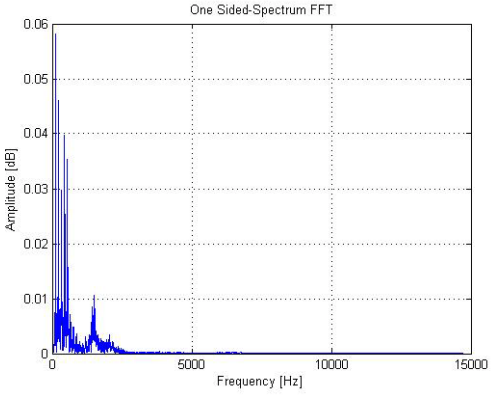




Obr. 5.12 Původní signál souhlásky „p“ Obr. 5.13 Signál souhlásky „p“ po FFT

Souhláska „r“:





Obr. 5.14 Původní signál souhlásky „r“ Obr. 5.15 Signál souhlásky „r“ po FFT

5.3 APLIKACE NEURONOVÉ SÍTĚ V PROSTŘEDÍ MATLAB

MATLAB je programové prostředí a skriptovací programovací jazyk pro vědeckotechnické numerické výpočty, modelování, návrhy algoritmů, počítačové simulace, analýzu a prezentaci dat, měření a zpracování signálů, návrhy řídicích a komunikačních systémů. Název MATLAB vznikl zkrácením slov MATrix LABoratory (volně přeloženo „laboratoř s maticemi“), což odpovídá skutečnosti, že klíčovou datovou strukturou při výpočtech v MATLABu jsou matice. [17]

5.3.1 Programování neuronové sítě

Vstupní množina vzorů, na které se neuronová síť učí, je tvořena daty získanými aplikací Fourierovy transformace na vstupní signál. Neuronovou síť typu SOM vytvoříme příkazem *newsom()* s příslušnými vstupními parametry.

```
%vstupni data
Input = load('vzorky.data');
Input=Input';

%vytvoreni neuronove site
PR = minmax(Input);                 %rozsah vstupnch hodnot
Di = [6 4];                         %pocet vrstev neuronu
TFCN = 'hextop';                    %topologie
DFCN = 'dist';                      %distancni fce
OLR = 0.9;                          %learning rate
OSTEPS = 1000;                      %pocet kroku
TLR = 0.02;                         %ladeni learning rate
TND = 1;                            %vzdalenost souseda
net = newsom(PR,Di,TFCN,DFCN,OLR,OSTEPS,TLR,TND);
```

Proměnná *TFCN* může nabývat různých parametrů podle zvolené topologie sítě *(hextop, gridtop, randtop)*. Proměnná *DFCN (distance function)*, která určuje vzdálenost mezi sousedními neurony, může nabývat parametrů *linkdist*, *dist* nebo *mandist*.

Do proměnné *net* se uloží data vytvořeného objektu neuronové sítě. Objekt *net* kóduje veškerá paradigmata neuronové sítě. Jsou zde uloženy informace o topologii, přenosových funkcích, chybových funkcích, hodnoty vah, prahů.

Dalším krokem je inicializace sítě, při které dochází k nastavení počátečních vah a prahů neuronů.

```
net = init(net);
```

Posledním krokem programování neuronové sítě je natrénování vytvořené neuronové sítě, které se provede příkazem *train()*. Před vlastním trénováním sítě zvolíme parametry trénování:

`net.trainParam.epochs` – počet trénovacích cyklů
`net.trainParam.goal` – nastavení hodnoty chybové funkce, při níž se ukončí trénování neuronové sítě
`net.trainParam.lr` – udává rychlost učení
`net.trainFcn` – volba trénovací funkce

```
%trenovani site
[net,tr,Y,E] = train(net,Input);
```

Vstupní hodnotou funkce *train()* je proměnná *Input*, ve které jsou uloženy trénovací vzory dat. Jejími výstupními proměnnými jsou *net* (název objektu trénované sítě), *tr* (průběh trénovacího procesu), *Y* (výstupy sítě), *E* (chyby sítě).

Po natrénování sítě zbývá ověřit, jak úspěšně se síť natrénovala. K simulaci slouží příkaz *sim()*. Jeho vstupními proměnnými jsou *net* a *Input*, výstupní proměnnou je *Output*.

Neodpovídá-li výsledek simulace svému vzoru, přenastaví se parametry trénování a postup trénování se opakuje, až se dosáhne optimálního výsledku.

6. ZÁVĚR

Bakalářská práce na téma „Rozpoznání jednotlivých písmen ve zvukovém záznamu s využitím SOM“ se sestává ze čtyř hlavních částí – klasické modely neuronových sítí, rozpoznávače přirozené řeči, analýza řečového signálu a vlastní navržený systém.

Podstatou první části je nahlédnutí do historie neuronových sítí. Obsahuje seznámení se základními pojmy z oblasti neuronových sítí, krátký popis práce a procesu jejich učení. Bylo vybráno pět nejpoužívanějších sítí, jejichž činnost zde byla popsána. Jsou to: Hopfieldova síť, Kohenova síť, Neuronová síť se zpětným šířením signálu, Síť pro interaktivní aktivaci a soutěžení a Neocognitron.

Druhá část představuje tři vybrané systémy pro rozpoznávání řeči pomocí neuronových sítí. U každého z nich je popsáno za jakým účelem byl vyvinut, jak pracuje a jeho úspěšnost je vyhodnocena pomocí konkrétního experimentu. Práce uvádí i přesná kritéria, při kterých experiment probíhal. Tímto způsobem jsou zpracovány: Network fusion, Subvokální komunikace a Generování cílů forward-backward šířením pravděpodobnosti.

Systém *network fusion* demonstruje, že spojením co největšího počtu podsítí dochází k výraznému snížení chybovosti systému pro rozpoznávání řeči.

Subvokální komunikace byla původně vyvinuta pro použití ve vesmíru, přesto nachází své uplatnění i na Zemi, například v hlučném prostředí nebo v ulehčení života tělesně postižených.

Vývoj a výzkum se soustředí také na hybridní systémy, jejichž největší předností je výrazné snížení chybovosti rozpoznávané řeči. Práce popisuje jen některé z nich.

Třetí část nazvaná „Analýza řečového signálu“ se zabývá charakteristikou lidské řeči a podává základní přehled systémů na její rozpoznávání. Poukazuje na rozdíly mezi jednotlivými systémy a jejich problémy.

Čtvrtá část pod názvem „Vlastní navržený systém“ se především soustředí na způsob získání analyzovaných dat, jejich zpracování pomocí algoritmu FFT a představuje několika vzorků vstupního signálu před a po aplikaci FFT. Závěr této části je věnován krátkému popisu programovacího prostředí MATLAB a konkrétním funkcím, pomocí nichž se neuronová síť programuje.

Zatímco vstupní data pro neuronovou síť se podařilo úspěšně zpracovat pro účely učení sítě, jak je znázorněno na příkladech frekvenčních spekter vzorků v kapitole 5.2.2, naprogramování neuronové sítě v prostředí MATLAB se nezdařilo do té míry, aby byl získán konkrétní výstup z algoritmu ve tvaru textového zápisu rozpoznaného písmene a uveden v této práci. Problémem bylo správné nastavení parametrů neuronové sítě.

Na rozpoznávání řeči se v posledních letech soustředí stále větší pozornost, neboť se dá očekávat její širší využití v mnoha oborech lidské činnosti.

LITERATURA

[1] KATAGIRI, S.: *Handbook of Neural Networks for Speech Processing*, Artech House INC, 2000, ISBN 0-89006-954-9.

[2] NEJEDLOVÁ, D.: *Fonetická transkripce češtiny pomocí třívrstvé neuronové sítě*, Výzkumná zpráva č. ISRN-TUL-KES-T-PZ-00-005-C1-CZ, TU Liberec, 2000.

[3] BERÁNEK, L. *Neuronové sítě* [online]. Dostupné z: <http://www.eamos.cz/amos/kat_inf/externi/kat_inf_34296/k0_uvod.pdf>

[4] STERGIOU, CH., SIGANOS, D. *Neural network* [online]. Dostupné z : < http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html>.

[5] DRÁBEK, O., SEIDL, P., TAUFER, I. Umělé neuronové sítě - základy teorie a aplikace. *CHEMagazín*, 2006, roč. XIV, č. 2, s. 33 – 36.

[6] WILSON, B. *Artificial Intelligence* [online]. 2006. Dostupné z: <http://www.cse.unsw.edu.au/~billw/cs9414/notes/ml/pdp /iac2005.html>.

[7] FÁBÍK, P., GAVLIAK, P. *Neokognitron* [online]. 2001. Dostupné z: <http://neuron-ai.tuke.sk/~fabik/Neocognitron.htm>.

[8] SOMMER, D. *Neuronale Netze* [online]. 1998. Dostupné z: <http://www.sund.de/netze/applets/BPN/bpn2/ochre.html>.

[9] VOJÁČEK, A. *Samoučící se neuronová síť* [online]. 2006. Dostupné z: <http://automatizace.hw.cz/mereni-a-regulace/ART244-samoucici-se--neuronova-sit--som-kohonenovy-mapy.html>.

[10] PAVELKA, T., Ekštein, K., Andrš, D.: *Hybridní rozpoznávač přirozené řeči pro český jazyk*, proc. Kognícia a umelý život 2005, Smolenice, Slovakia, 2005. Dostupný z: <http://hilbert.chtf.stuba.sk/KUZV/download/kuzv-pavelka-ekstein-andrs.pdf>

[11] BUHRKE, R. E, LOCICERO, J. L.: *Speech Recognition with Neural Network and Network Fusion*, IEEE, s. 157 – 160, 1991.

[12] MENDES, J. AG, ROBSON, R. R., LABIDI, S., BARROS, K. A. Subvocal Speech Recognition Based on EMG signal Using Independent Component Analysis and Neural Network MLP. In *Proceedings IEEE Congres of Signal and Image Processing*, s. 221 - 224, 2008.

[13] YONGHONG, Y., FANTY, M., COLE, R. Speech Recognition Using Neural Networks with Forward-Backward Probability Generated Targets. In *Proceedings IEEE International Conference on Acustics, Speech and Signal Processing*, s. 3241 – 3244, 1997.

[14] JURČÍČEK, F. *Dekodér systému pro rozpoznávání souvislé mluvené řeči s velkým slovníkem (LVCSR) s n-gramovým jazykovým modelem.* Diplomová práce. Plzeň: ZČU, 2003.

[15] MELICHAR, J.,STYBLÍK, V.: *Český jazyk*, Nakladatelství Fortuna, 1994, ISBN 80-7168-130-X.

[16] FARSKÝ, J. *Rozšíření funkcí demonstračního robota Festík akustické vstupy a výstupy.* Semestrální projekt. Praha, ČVUT, 2002.

[17] WIKIPEDIE. Dostupné z: < http://cs.wikipedia.org/wiki/MATLAB>

PŘÍLOHA

Pojmy z oblasti neuronových sítí

Adaline

Klasická umělá neuronová síť perceptronovského typu s binárními výkonnými prvky. Jejich váhy jsou nastavitelné a učení probíhá tzv. delta pravidlem.

Adaptace

Schopnost umělé neuronové sítě přizpůsobit se.

Architektura

Struktura sítě výkonných prvků, jejich vzájemné propojení.

Bázový prvek

Jeden z prvků umělé neuronové sítě, který je stále aktivní. Jeho výstup se přivádí do všech ostatních výkonných prvků jako jejich práh.

Cílový vektor

Žádaný výstupní vektor patřící k nějakému vstupnímu vektoru. Tento vektor musí být při učení s učitelem znám.

Delta pravidlo

Pravidlo pro učení s učitelem, u kterého se změnou vah dosahuje stále se zmenšujícího rozdílu mezi žádanou a skutečně dosaženou hodnotou výstupu.

Dopředná (nerekurzivní) síť

Moderní vícevrstvá síť. Je v ní jednoznačně definován informační tok. V takové síti neexistuji spoje mezi neurony z vyšších vrstev zpět do vrstev nižších, dokonce ani spojení mezi neurony v téže vrstvě.

Genetické algoritmy

Jsou inspirovány přirozeným chováním přírody, v níž probíhá evoluční vývoj.

Kompetiční učení

Učící pravidlo, ve kterém si výkonné prvky při předkládání vstupních vzoru vzájemně konkurují. Váhy se pak mohou měnit pouze u vítězného neuronu.

Madaline

Je o jednu vrstvu rozšířená síť Adaline. Má zvláštní metodu učení, protože na rozdíl od Adaline obsahuje jednu skrytou vrstvu.

Neuron

Buňka nervového systému. Neuron je anatomicky i funkčně základním stavebním kamenem nervového systému. Posloužil jako vzor pro výkonný prvek v umělých neuronových sítích.

Perceptron

Jednoduchá dopředná síť bez skrytých vrstev. To znamená, že jen jednu vrstvu této sítě lze učit. Klasickou hranicí schopností perceptronu je XOR-problém.

Práh

Hodnota, kterou musí součet všech vážených vstupů neuronu překročit, aby se stal aktivním.

Přeučování

Učící proces, ve kterém se maže jistý počet vah. V kontrastu k normálnímu učení už při přeučování síť jistý objem vědomostí obsahovala. Muže též jít o stav, kdy síť už překonala zenit svých možností a začíná chybovat.

Samoorganizace

Schopnost neuronové sítě učením bez učitele přizpůsobit své chování k vyřešení daného problému.

Schopnost asociace

Vlastnost neuronové sítě odhalit podobnosti mezi naučenými vzory a vstupními daty.

Synapse

Místo styku mezi dvěma neuronovými buňkami v organismu. Během učení se jeho parametry mění.

Šum

Náhodné změny některých informačních jednotek, které dohromady představují vstupní vzor. Neuronové sítě mají schopnost rozpoznat i zašuměné vstupní vzory.

Tolerance k chybám

Schopnost neuronové sítě odpovědět i na vzor, který se poněkud liší od toho, který byl součástí trénovací množiny.

Topologie

Popisuje druh a počet výkonných prvků sítě a její strukturu.

Učení

Přizpůsobování nebo adaptace neuronové sítě daným požadavkům. Váhy na spojích mezi jednotlivými výkonnými prvky sítě se mění podle nějakého učícího algoritmu.

Učící fáze

Časový interval, během kterého se podle nějakého učícího algoritmu mění parametry sítě a tyto se do sítě nahrávají.

Učící pravidlo (algoritmus)

Předpis, který udává, jak se budou síti předkládat vzory k učení a jak se budou vypočítávat změny vah.

Váha

Hodnotou vyjádřená míra vazby mezi dvěma spojenými výkonnými prvky. Jejím prostřednictvím se v síti uchovávají informace. Paměť sítě představují právě tyto váhy, resp. jejich velikosti.

Vrstva

Základní komponenta architektury neuronové sítě. Vrstvu tvoří jistý počet stejných buněk majících v síťové struktuře identickou funkci.

Zobecňování

Schopnost neuronové sítě na základě naučených vzorů odpovědět i na vzor, který nebyl součástí učící množiny.

Zpětná vazba

Zvláštní propojení výkonných prvků podobné např. kruhu, kdy se informační tok znovu vrací ke svému výchozímu bodu.